BOEING SCIENTIFIC RESEARCH LABORATORIES DOCUMENT D1-82-0955

BOEING

SCIENTIFIC RESEARCH LABORATORIES

Random Variables with Independent Binary Digits

George Marsaglia

MATHEMATICS RESEARCH JANUARY 1970

15

D1-82-0955

RANDOM VARIABLES WITH INDEPENDENT BINARY DIGITS

by

George Marsaglia

Mathematical Note No. 641

Mathematics Research Laboratory

BOEING SCIENTIFIC RESEARCH LABORATORIES

January 1970

Abstract

Let $X = .b_1b_2b_3\ldots$ be a random variable with independent binary digits b_n taking values 0 or 1 with probabilities p_n and q_n. When does X have a density function? A continuous density function? A singular distribution? This note proves that the distribution of X is singular if and only if the tail of the series $\sum[\log(p_n/q_n)]^2$ diverges, and that X has a density that is positive on some interval if and only if $\log(p_n/q_n)$ is a geometric sequence with ratio $\frac{1}{2}$ for n greater than some k, and in that case the fractional part of 2^kX has an exponential density (increasing or decreasing with the uniform density a special case). It gives a sufficient condition for X to have a density, ($\sum\log[2\ \max(p_n,q_n)]$ converges), but unless the tail of the sequence $\log(p_n/q_n)$ is geometric, ratio $\frac{1}{2}$, the density is a weird one that vanishes at least once in every interval.

1. Introduction

It is well known that one can construct a uniform random variable by choosing the binary digits with successive flips of a good coin, ($p = \frac{1}{2}$). Such considerations date back to the beginnings of probability theory--indeed, to the development of measure and integration theory.

For bad coins, $0 < p < 1$, $p \neq \frac{1}{2}$, the resulting number has a distribution that is continuous but singular, as its possible values form a set of Lebesgue measure zero. This note is concerned with the case where the binary digits are independent, but not identically distributed. Are there any interesting random variables that arise from this situation? It turns out that there are some interesting singular distributions, and that there are distributions which have densities. This note will show that there is essentially one conventional type density that can arise this way--the exponential (increasing or decreasing, with the uniform a special case), if by conventional we mean a density that is positive on some interval. We will also find necessary and sufficient conditions that the distribution be singular. The conclusion is that independent binary digits lead to one of four possibilities:

1) A singular distribution. This happens if and only if the series $\sum_{n=m}^{\infty} \log^2(p_n/q_n)$ diverges for every m. A special, but uninteresting, case is the discrete distribution arising when $p_n q_n = 0$ for all suitably large n. Some interesting continuous but singular distributions arise. See Section 3.

2) A piecewise-exponential density, with pieces equally spaced and of similar shape. This happens only when the tail of the sequence

$\log(p_n/q_n)$ is geometric with ratio $\frac{1}{2}$. It is the only way to get a distribution which has a positive derivative on some interval. See Section 2.

3) A distribution with a density, but a strange density that vanishes at least once on very interval. This will happen if $\sum \log[2 \max(p_n,q_n)]$ converges and $\log(p_n/q_n)$ is not geometric with ratio $\frac{1}{2}$. Section 4.

4) A distribution with a Lebesgue decomposition having both absolutely continuous and singular parts. We conjecture that this class is empty, but have not been able to rule out the possibility for cases where $\sum \log[2 \max(p_n,q_n)]$ diverges but $\sum \log^2(p_n/q_n)$ converges.

2. Distributions with Reasonable Densities

The most interesting case seems to be the assignment of probabilities to the bits so that the resulting random variable has a conventional density function. We will show that there is essentially only one way to do this. The general statement is Theorem 2 below, but by shifting the binary decimal point to the right far enough we may assume we are dealing only with the fractional part taking values on the unit interval, and in that case we can formulate the basic requirement as follows:

Theorem 1. *If* X *is a random variable on the unit interval with independent binary digits,*

$$X = .b_1b_2b_3\ldots = \sum_{i=1}^{\infty} b_i 2^{-i},$$

b's *independently* 0 *or* 1, *and if the distribution function of* X, *say* F(x), *has a positive derivative at* $\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \ldots$ *that is,* F'(.1), F'(.01), F'(.001), ... *all exist and are positive, then* X *has an exponential distribution with density*

$$f(x) = \frac{\beta e^{\beta x}}{e^{\beta} - 1}, \quad 0 < x < 1, \quad -\infty < \beta < \infty$$

and the probabilities for the bits of X *are given by the formula*

$$P[b_i = 0] = \frac{1}{1 + e^{\beta/2^i}}$$

$$P[b_i = 1] = \frac{e^{\beta/2^i}}{1 + e^{\beta/2^i}}$$

Proof: Represent the probabilities for the bits of X as follows:

$$P[b_i = 0] = \frac{1}{1 + e^{c_i}}$$

$$P[b_i = 1] = \frac{e^{c_i}}{1 + e^{c_i}}$$

We will get a relation between the c's by representing F'(.1) as the limit of two sequences of difference quotients:

(1) $$\frac{F(.101) - F(.1)}{2^{-3}}, \quad \frac{F(.1001) - F(.1)}{2^{-4}}, \quad \frac{F(.10001) - F(.1)}{2^{-5}}, \ldots$$

and

(2) $$\frac{F(.1) - F(.01)}{2^{-2}}, \quad \frac{F(.1) - F(.011)}{2^{-3}}, \quad \frac{F(.1) - F(.0111)}{2^{-4}}, \ldots$$

The general term of sequence (1) has the form

$$\frac{e^{c_1}}{\prod_{i=1}^{n}\left(\frac{1+e^{c_i}}{2}\right)}$$

and the general term of sequence (2) has the form

$$\frac{e^{c_2+c_3+\ldots+c_n}}{\prod_{i=1}^{n}\left(\frac{1+e^{c_i}}{2}\right)} .$$

Since $F'(.1)$ exists and is positive, we conclude that the infinite product converges, that

$$F'(.1) = \frac{e^{c_1}}{\prod_{1}^{\infty}\left(\frac{1+e^{c_i}}{2}\right)} = \frac{e^{c_2+c_3+\ldots}}{\prod_{1}^{\infty}\left(\frac{1+e^{c_i}}{2}\right)}$$

and hence $c_1 = c_2 + c_3 + \ldots$.

A similar argument on $F'(.01)$ shows that $c_2 = c_3 + c_4 + \ldots$; on $F'(.001)$ shows $c_3 = c_4 + c_5 + \ldots$ and thus we conclude that there is a β such that $c_1 = \beta/2$, $c_2 = \beta/4$, $c_3 = \beta/8$, $\ldots$. This provides the formulas for the bits of X as given in the theorem. We still must show that if the bits of X are given by those probabilities, then X has an exponential density with parameter β. (Note that β can be either positive or negative, with $\beta = 0$ giving the uniform density.)

Writing

$$X = \frac{b_1}{2^1} + \frac{b_2}{2^2} + \frac{b_3}{2^3} + \ldots$$

we express the characteristic function of X as an infinite product

$$\prod_{k=1}^{\infty}\left[\frac{1+e^{\frac{\beta+it}{2^k}}}{1+e^{\beta/2^k}}\right] .$$

Using the relation

$$1 - e^{z} = (1+e^{z/2})(1-e^{z/2}) = (1+e^{z/2})(1+e^{z/4})(1-e^{z/4}) = \ldots$$

we have

$$\prod_{k=1}^{n-1}(1+e^{z/2^k}) = \frac{1 - e^z}{1 - e^{z/2^n}}$$

and hence

$$\prod_{k=1}^{n-1}\left[\frac{1+e^{z/2^k}}{1+e^{\beta/2^k}}\right] = \left[\frac{1-e^z}{1-e^\beta}\right] \left[\frac{1-e^{\beta/2^n}}{1-e^{z/2^n}}\right]$$

The right side converges to

$$\left[\frac{1-e^z}{1-e^\beta}\right] \left[\frac{\beta}{z}\right]$$

which, with $z = \beta + it$, is the characteristic function of X with density $\gamma e^{\beta x}$ on $0 < x < 1$.

To get a more general theorem, we note that if Y has a positive density on some interval $a < y < b$ then there are integers r and k such that Y has a positive density on a subinterval of the form

$$a \le \frac{r}{2^k} < y < \frac{r+1}{2^k} \le b$$

and hence the fractional part of $2^k Y$ has a density on the unit interval; the above theorem applies, and we have:

Theorem 2. *If Y is a random variable with independent binary digits, and if Y has a distribution which has a positive derivative on some interval, $F'(y) > 0$, $a < y < b$, then Y may be scaled by a power of 2, that is, its binary decimal may be relocated, so that its fractional part is exponentially distributed with density*

$$\frac{\beta e^{\beta x}}{e^\beta - 1} \qquad 0 < x < 1$$

(For some β, $-\infty < \beta < \infty$, with the uniform density corresponding to $\beta = 0$).

In other words, Y *may be represented in the form* $2^{k(M+X)}$ *where* M *is a random integer, independent of the fractional part* X *having density* $\gamma e^{\beta x}$, $0 < x < 1$.

3. Singular Distributions

Let $p_n = P[b_n = 0]$. If some subsequence of the p's converges to a value other than $\frac{1}{2}$, then the distribution of $X = .b_1 b_2 b_3 \ldots$ will be singular, for its possible values will form a set of Lebesgue measure 0. It is easy to get many representations of singular variates in this way. In particular, one can get two singular distributions whose convolution has an exponential or a uniform distribution, by writing

$$X_1 = .0b_2 0b_4 0b_6 0 \ldots$$
$$X_2 = .b_1 0b_3 0b_5 0b_7 \ldots$$

where the b's take values with probabilities given by Theorem 1. Then $X_1 + X_2$ has an exponential density $\gamma e^{\beta x}$ on $0 < x < 1$.

To get two singular distributions whose convolution is the ordinary exponential density $\alpha e^{-\alpha x}$, $0 < x < \infty$, write

$$X = \ldots d_3 d_2 d_1 d_0 . d_{-1} d_{-2} d_{-3} \ldots$$

where

$$P[d_k = 0] = \frac{1}{1 + e^{\alpha 2^k}}$$

and

$$X_1 = \ldots d_3 0 d_1 0 . d_{-1} 0 d_{-3} 0 d_{-5} \ldots$$
$$X_2 = \ldots d_4 0 d_2 0 d_0 . 0 d_{-2} 0 d_{-4} 0 \ldots$$

We now turn to the general question of when the distribution of $X = .b_1b_2b_3\ldots$ is singular, where b_i takes values 0 or 1 with probabilities p_n and q_n. If p_n does not converge to $\frac{1}{2}$ then the distribution of X will be singular, but what happens when $p_n \to \frac{1}{2}$ but not according to the formulas of Theorem 1? We know that X cannot have a continuous density but if $p_n \to \frac{1}{2}$ very very quickly we might expect that X will have some kind of a density, though a weird one. The answer is yes, there is always a density if $p_n \to \frac{1}{2}$ quickly enough. Alternatively, we will prove that the distribution of X is singular (including discrete) if and only if $\sum_{n=m}^{\infty} \log^2(p_n/q_n) = \infty$ for all positive integers m.

To prove this result we need a preliminary lemma which gives a formula for $F'(x)$ when it exists:

Lemma 1. *Let* $X = .b_1b_2b_3\ldots$ *have independent binary digits with*

$$P[b_n=0] = p_n,\ P[b_n=1] = q_n\ .$$

Let F *be the distribution function of* X. *If* F' *exists at* $v = .v_1v_2v_3\ldots$ *then*

$$F'(v) = [2g_1(v_1)][2g_2(v_2)][2g_3(v_3)]\ \ldots \tag{3}$$

where $g_n(0) = p_n$ *and* $g_n(1) = q_n$.

Proof: Since $F'(v)$ exists it can be represented

$$F'(v) = \lim_{\substack{s,t\to v \\ s<v<t}} \frac{F(t) - F(s)}{t - s}$$

and we may write

$$F'(.v_1v_2v_3\ldots) = \lim_{n\to\infty} 2^n[F(.v_1v_2\ldots v_n+2^{-n}) - F(.v_1v_2\ldots v_n)].$$

Then (3) follows from the fact that the expression in brackets is $g_1(v_1)g_2(v_2)\ldots g_n(v_n)$.

Theorem 3. *Let* $X = .b_1b_2b_3\ldots$ *have independent binary digits with* b_n *taking values* 0 *or* 1 *with probabilities* p_n *and* q_n. *In order that* X *have a singular distribution function (derivative equal zero almost everywhere) it is necessary and sufficient that for every positive integer* m,

$$\sum_{n=m}^{\infty} [\log(p_n/q_n)]^2 = \infty.$$

Proof: Let F be the distribution function. It has a finite derivative almost everywhere. Thus from Lemma 1, for almost all $x = .x_1x_2x_3\ldots$ we have F' expressed as an infinite product:

$$F'(.x_1x_2x_3\ldots) = [2g_1(x_1)][2g_2(x_2)][2g_3(x_3)]\ \ldots\ , \tag{4}$$

where $g_n(0) = p_n$ and $g_n(1) = q_n$.

Another interpretation of the fact that F has a derivative almost everywhere is to say that if $.v_1v_2v_3\ldots$ is chosen at random with the v's independently 0 or 1 with probability $\frac{1}{2}$, then with probability 1 the product in (3) converges to a (possibly zero) constant. According to a standard theorem on infinite products, (see, e.g., Knopp [2], p. 223), the tail of the product in (4) converges to a non-zero constant e^L if and only if for some m,

$$\sum_{n=m}^{\infty} \log[2g_n(x_n)] = L\ .$$

Thus the question of whether F' is positive almost everywhere or zero almost everywhere hinges on the convergence of the random series

$$\sum_{n=m}^{\infty} \log[2g_n(v_n)],$$

where the v's take values 0 or 1 with probability $\frac{1}{2}$.

We apply the three series theorem (see, e.g., Fisz [1], p. 248) after computing

$$E\{\log[2g_n(v_n)]\} = \frac{1}{2}\log(4p_nq_n), \text{ Variance} = \frac{1}{4}[\log(p_n/q_n)]^2.$$

If the tail of $\sum \log^2(p_n/q_n)$ converges, so does the tail of $-\sum\log(4p_nq_n)$, for the terms of the former dominate the terms of the latter. (Write $2p_n = 1 + t$, $2q_n = 1 - t$ then note that $\log^2[(1+t)/(1-t)] + \log(1-t^2) \geq 0$ for $-1 < t < 1$.)

Applying the three series theorem to the random series $\sum\log(2g_n(v_n))$ and interpreting the result in terms of the infinite product (4) we conclude:

If $\sum_{n=m}^{\infty} \log^2(p_n/q_n)$ converges for some m, then $F'(x) > 0$ for almost all x; if it diverges for all m, then $F'(x) = 0$ for almost all x and hence F is singular.

4. Absolute Continuity

To recapitulate: We know that F is singular if and only if the tail of $\sum\log^2(p_n/q_n)$ diverges, and hence that F has an absolutely continuous component in its Lebesgue decomposition if and only if the tail of $\sum\log^2(p_n/q_n)$ converges. But does the convergence of the tail of $\sum\log^2(p_n/q_n)$ imply that F is purely absolutely continuous, or can it have both an a.c. and a singular part? The question is open, but the following theorem shows that if $p_n \to \frac{1}{2}$ fast enough, then F will have no singular part.

Theorem 4. *Let* $X = .b_1b_2b_3\ldots$ *have independent binary digits with* b_n *taking values* 0 *or* 1 *with probabilities* p_n *and* q_n. *If the series*

$$\sum_{i=1}^{\infty} \log[2 \max(p_i,q_i)] = \sum_{i=1}^{\infty} \log[1+|2p_i-1|]$$

converges then the distribution of X *is absolutely continuous.*

Proof: Let

$$\prod_{i=1}^{\infty}[2 \max(p_i,q_i)] = M,$$

where $0 < M < \infty$ because the series of logarithms converges. We will show that F is absolutely continuous by showing that the probability of an interval is less than M times the length of the interval:

$$F(y) - F(x) \le (y-x)M, \quad x < y. \tag{5}$$

In fact, it suffices to show that (5) holds if x and y are in the dense set of terminating binary decimals, and in that case we may write, with $g_i(0) = p_i$ and $g_i(1) = q_i$,

$$\begin{aligned} F(.x_1x_2\ldots x_n+2^{-n}) - F(.x_1x_2\ldots x_n) &= g_1(x_1)g_2(x_2)\ldots g_n(x_n) \\ &\le 2^{-n}\prod_{i=1}^{n}\max(2p_i,2q_i) \\ &\le 2^{-n}M. \end{aligned}$$

5. A Conjecture

There seems to be one question that remains, which we put in the form of a conjecture: *If* $X = .b_1b_2b_3\ldots$ *has independent binary digits, then the distribution of* X *is either purely discrete, purely absolutely continuous, or purely continuous and singular.*

It is easy to show that it is purely discrete if and only if $p_n q_n = 0$ for all large n, and it is purely singular, including discrete, if and only if the tail of $\sum \log^2(p_n/q_n)$ diverges. But what about when the tail converges, but not fast enough for $\sum \log[2 \max(p_n, q_n)]$ to converge? The resulting F has a positive derivative almost everywhere, but can the Lebesgue decomposition of F have both an absolutely continuous and a singular part?

6. Acknowledgement

I would like to thank Albert W. Marshall and Gordon B. Crawford for stimulating discussions on the problem of independent binary digits; in particular, A. W. Marshall suggested the sufficient condition of Theorem 4.

References

[1] Fisz, Marek. *Probability Theory and Mathematical Statistics*, Third Edition, John Wiley and Sons, New York 1963.

[2] Knopp, Konrad. *Theory and Application of Infinite Series*, Blackie and Son, London 1948.